A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP6345.

Cadmium, Smoking, and Human Blood DNA Methylation Profiles in Adults from the Strong Heart Study

Arce Domingo-Relloso,[1,13,14] Angela L. Riffo-Campos,[2] Karin Haack,[3] Pilar Rentero-Garrido,[4,5] Christine Ladd-Acosta,[6] Daniele M. Fallin,[5,6] Wan Yee Tang,[7] Miguel Herreros-Martinez,[8] Juan R. Gonzalez,[10,11,12] Anne K. Bozack,[1] Shelley A. Cole,[3] Ana Navas-Acien,[1] and Maria Tellez-Plaza[7,9,13]

[1]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, New York, USA
[2]Department of Pathology, Universidad de La Frontera, Temuco, Chile
[3]Population Health Program, Texas Biomedical Research Institute, San Antonio, Texas, USA
[4]Precision Medicine Unit, Institute for Biomedical Research INCLIVA, Valencia, Spain
[5]Department of Mental Health, Johns Hopkins University, Baltimore, Maryland, USA
[6]Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland, USA
[7]Department of Environmental Health and Engineering, Johns Hopkins University, Baltimore, Maryland, USA
[8]Bioinformatics Unit, Institute for Biomedical Research INCLIVA, Valencia, Spain
[9]Area of Cardiometabolic and Renal Risk, Institute for Biomedical Research INCLIVA, Valencia, Spain
[10]Institute for Global Health (ISGlobal), Barcelona, Spain
[11]Universitat Pompeu Fabra (UPF), Barcelona, Spain
[12]CIBER en Epidemiología y Salud Pública (CIBERESP), National Center of Epidemiology, Carlos III Health Institute, Madrid, Spain
[13]Department of Chronic Diseases Epidemiology, National Center of Epidemiology, Carlos III Health Institute, Madrid, Spain
[14]Department of Statistics and Operations Research, University of Valencia, Spain

Background: The epigenetic effects of individual environmental toxicants in tobacco remain largely unexplored. Cadmium (Cd) has been associated with smoking-related health effects, and its concentration in tobacco smoke is higher in comparison with other metals.

Objectives: We studied the association of Cd and smoking exposures with human blood DNA methylation (DNAm) profiles. We also evaluated the implication of findings to relevant methylation pathways and the potential contribution of Cd exposure from smoking to explain the association between smoking and site-specific DNAm.

Methods: We conducted an epigenome-wide association study of urine Cd and self-reported smoking (current and former vs. never, and cumulative smoking dose) with blood DNAm in 790,026 CpGs (methylation sites) measured with the Illumina Infinium Human MethylationEPIC (Illumina Inc.) platform in 2,325 adults 45–74 years of age who participated in the Strong Heart Study in 1989–1991. In a mediation analysis, we estimated the amount of change in DNAm associated with smoking that can be independently attributed to increases in urine Cd concentrations from smoking. We also conducted enrichment analyses and *in silico* protein–protein interaction networks to explore the biological relevance of the findings.

Results: At a false discovery rate (FDR)-corrected level of 0.05, we found 6 differentially methylated positions (DMPs) for Cd; 288 and 17, respectively, for current and former smoking status; and 77 for cigarette pack-years. Enrichment analyses of these DMPs displayed enrichment of 58 and 6 Gene Ontology and Kyoto Encyclopedia of Genes and Genomes gene sets, respectively, including biological pathways for cancer and cardiovascular disease. In *in silico* protein-to-protein networks, we observed key proteins in DNAm pathways directly and indirectly connected to Cd- and smoking-DMPs. Among DMPs that were significant for both Cd and current smoking (annotated to PRSS23, AHRR, F2RL3, RARA, and 2q37.1), we found statistically significant contributions of Cd to smoking-related DNAm.

Conclusions: Beyond replicating well-known smoking epigenetic signatures, we found novel DMPs related to smoking. Moreover, increases in smoking-related Cd exposure were associated with differential DNAm. Our integrative analysis supports a biological link for Cd and smoking-associated health effects, including the possibility that Cd is partly responsible for smoking toxicity through epigenetic changes. https://doi.org/10.1289/EHP6345

Introduction

Numerous studies have evaluated the association of tobacco smoke, a complex mixture of compounds, with several epigenetic marks, in particular with blood DNA methylation (DNAm) (Harlid et al. 2014; Joubert et al. 2012; Shenker et al. 2013; Sun et al. 2013; Tsaprouni et al. 2014; Wan et al. 2012; Zeilinger et al. 2013). For instance, DNAm of multiple CpGs (methylation sites) in the *AHRR* [Aryl Hydrocarbon Receptor Repressor, which mediates dioxin toxicity and is involved in cell growth and differentiation (Haarmann-Stemmann et al. 2007)] and *F2RL3* [coagulation factor II receptor-like 3, also known as PAR-4, which plays a role in blood coagulation, inflammation, and response to pain, (Heuberger and Schuepbach 2019)] genes has been associated with former and current smoking status, cumulative smoking, and shorter time since smoking cessation in several studies (Bojesen et al. 2017; Breitling et al. 2012; Fasanelli et al. 2015; Gao et al. 2015; Joehanes et al. 2016; Reynolds et al. 2015; Zhang et al. 2013b). In addition to *AHRR* and *F2RL3*, other genes have also been consistently and specifically associated with smoking in methylome-wide epidemiological studies investigating as many as 450,000 CpGs (Gao et al. 2015; Joehanes et al. 2016). The potential epigenetic effects of individual environmental toxicants in tobacco, however, remain largely unexplored.

Cadmium (Cd) is one of the main metals found in cigarettes (Nordberg et al. 2015; Tellez-Plaza et al. 2012). Its concentrations are extremely high in comparison with those from other metals such as arsenic or lead (Chiba and Masironi 1992). It is one of the most relevant heavy metals in terms of potentially explaining adverse health effects of smoking because it has been associated with tobacco-related health outcomes, including cardiovascular

Address correspondence to M. Tellez-Plaza, Department of Chronic Diseases Epidemiology, National Center for Epidemiology, Carlos III Health Institute, Melchor Fernandez Almagro, 5, 28029 Madrid, Spain. Email: m.tellez@isciii.es

Supplemental Material is available online (https://doi.org/10.1289/EHP6345).

The authors declare they have no actual or potential competing financial interests.

Received 5 October 2019; Revised 18 March 2020; Accepted 22 April 2020; Published 2 June 2020.

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

disease (CVD) and cancer (Agarwal et al. 2011; Andersson et al. 2018; García-Esquinas et al. 2014; Navas-Acien et al. 2004). Increased exposure to Cd has been associated with differences in global DNAm in human (Ruiz-Hernandez et al. 2015) and animal (Nica et al. 2017; Šrut et al. 2017) studies, possibly related to Cd-induced alterations of DNA methyltransferases (Benbrahim-Tallaa et al. 2007; Jiang et al. 2008; Poirier and Vlasova 2002; Takiguchi et al. 2003; Yuan et al. 2013) and ten-eleven translocation (TET) (Ravichandran 2017; Xiong et al. 2017) enzymes activity. Some reports have suggested that TETs activity is influenced by oxidative stress (Delatte et al. 2015). Cd has been related to oxidative stress (Chia et al. 2011; Domingo-Relloso et al. 2019; Lee et al. 2009; Moitra et al. 2014; Srivastava et al. 2014) and could interfere in the DNAm pathways by promoting redox unbalance. We hypothesized that Cd exposure from tobacco could induce changes in genomic DNAm and mediate well-known smoking-related methylation alterations. *In vitro* and *in vivo* studies have shown both gene-specific hypermethylation and hypomethylation in relation to Cd exposure (Hossain et al. 2012; Sanders et al. 2014; Vidal et al. 2015; Virani et al. 2016; Zhang et al. 2013a). The number of studies evaluating the association of Cd exposure with differences in CpG-specific DNAm in humans, however, is not sufficient (Ruiz-Hernandez et al. 2015).

Participants of the Strong Heart Study (SHS), a prospective cohort that includes 13 American Indian tribes, have substantially higher concentrations of urinary Cd in comparison with other large studies, such as the Multi-Ethnic Study of Atherosclerosis (MESA) and National Health and Nutrition Examination Survey (NHANES) (Deen et al. 2017). Cd exposure in the SHS was previously associated with smoking-related health effects, including total, lung, and pancreas cancer mortality (García-Esquinas et al. 2014); incident peripheral arterial disease (Tellez-Plaza et al. 2013b); and incident CVD (Tellez-Plaza et al. 2013a). The objective of this research was to conduct an epigenome-wide association study (EWAS) of Cd and smoking exposure with differences in human blood DNAm profiles. As a secondary objective, we assessed the implication of findings to relevant DNAm and demethylation pathways and evaluated whether the smoking-related increase in urine Cd levels can explain the association between smoking and genomic DNAm profiles. To achieve these aims, we evaluated the association of urinary Cd and smoking (self-reported active and former smoking, and cumulative smoking dose) with blood DNAm in 790,026 CpGs determined with the Infinium Methylation EPIC array in the SHS. We used a causal inference mediation approach (Pearl 2012) to estimate how much of the DNAm changes associated with smoking can be independently attributed to smoking-related increases in Cd exposure, as measured in urine (an established biomarker of Cd internal dose). In addition, the association analyses were followed by a gene set and molecular regulatory elements enrichment and *in silico* evaluation of relevant biological pathways for Cd and smoking influences in DNAm metabolism.

Methods

Study Population

The SHS is a prospective cohort study funded by the National Heart, Lung, and Blood Institute to investigate CVDs and its risk factors in American Indian adults (Lee et al. 1990). In 1989–1991, a total of 4,549 men and women ages 45–75 y who were members of 13 tribes based in Arizona, Oklahoma, North Dakota, and South Dakota accepted invitations to participate. All individuals were asked to fast overnight (8 h or more). Participants without sufficient urine for metal analyses were excluded ($n = 576$) (Figure S1). Due to tribal request, samples from one of the tribes were not selected for DNAm analyses, leaving 3,516 participants. Among them, participants who were free of CVD and not missing urinary metals or other variables of interest at baseline (1989–1991) were eligible for blood DNAm analyses ($n = 3{,}106$). Sufficient blood was available for DNAm analysis in 2,351 participants. After laboratory analyses, we removed data from 18 individuals without classical bimodal distribution in DNAm levels and from 8 individuals with low median intensity levels, leaving a total of 2,325 participants for this study. These participants were similar to those eligible in sociodemographic and anthropometric characteristics (Table S1).

Participant Characteristics and Self-Reported Smoking

Trained and certified nurses and medical examiners collected information on sociodemographic factors (age, sex, study region, education level), medical history, smoking status (never, former, current), and cumulative smoking dose (cigarette pack-years) in a personal interview. Participants having smoked ≥ 100 cigarettes in their lifetime and smoking at the time of the interview were considered current smokers. Noncurrent smokers who had smoked > 100 cigarettes in their lifetime were classified as former smokers. Cigarette pack-years were calculated as the number of 20-cigarette packs smoked per day times the number of years the person smoked, with zero assigned to never smokers. The trained nurses and medical examiners conducted a physical exam, including anthropometric measures [height and weight to measure body mass index (BMI)], and collected fasting blood and spot urine samples. Plasma creatinine was measured by an alkaline picrate rate method (Lee et al. 1990). We calculated estimated glomerular filtration rate (eGFR) from recalibrated plasma creatinine, age, and sex using the Chronic Kidney Disease – Epidemiology Collaboration formula (Levey et al. 2009).

Urinary Cadmium Determinations

Morning spot urine samples were collected in polypropylene tubes, frozen within 1 to 2 h of collection, shipped buried in dry ice, and stored at -70°C in the Penn Medical Laboratory, MedStar Research Institute, Washington, DC, USA. In 2009, urine samples were shipped to the Trace Metals Laboratory at Graz University, Austria, to measure Cd and other metals using inductively coupled plasma–mass spectrometry (ICP-MS) (Agilent 7700x ICP–MS; Agilent Technologies). The limit of detection (LOD) for urine Cd was 0.015 μg/L. In one participant below the LOD, the Cd concentration was imputed as the LOD divided by the square root of two. Urine Cd concentrations were corrected for molybdenum oxide interference. Other laboratory details and extensive quality control/quality assurance have been published (Scheer et al. 2012). To account for urine dilution, urine Cd concentrations were expressed in micrograms per gram of urine creatinine. Urine creatinine was measured at the Laboratory of the National Institute of Diabetes and Digestive and Kidney Disease, Epidemiology and Clinical Research Branch (Phoenix, Arizona, USA) by an alkaline picrate rate method.

Microarray DNAm Determinations

Buffy coats from fasting blood samples were collected in 1989–1991, at the same time as the collection of urinary samples for Cd measures and of physical examination and interview for the assessment of baseline smoking status and sociodemographic variables. Biological specimens were stored at -70°C. DNA from white blood cells was extracted and stored at the Penn Medical Laboratory, MedStar Health Research Institute under a strict quality-control system. In 2015, blood DNA was shipped to the analytical laboratory at the Texas Biomedical Research Institute for DNAm analysis. DNA was bisulfite-converted with the EZ DNAm kit (Zymo Research) according to the manufacturer's instructions. Bisulfite-converted

DNA was measured using the Illumina MethylationEPIC BeadChip (850K), which provides a measure of DNAm at a single nucleotide resolution at >850,000 CpGs. Samples were randomized across and within plates to remove potential batch artifacts and confounding effects, and replicate and across-plate control samples were included on every plate. All the preprocessing was conducted using R version 3.6.1. Data were read in six different batches (of ~400 individuals each) and combined using the R package minfi (version 1.18.4). CpGs with a *p*-detection value greater than 0.01 in more than 5% of the individuals (6,159 CpGs) were removed. Single sample normalization was conducted using the preprocessNoob function in R package minfi (Fortin et al. 2017; Triche et al. 2013), which includes a background correction with dye-bias normalization for Illumina Infinium methylation arrays. As a result of these preprocessing preliminary analyses, we had data from 2,325 individuals and 860,079 CpGs. Cross-hybridizing probes, sex chromosomes, and SNP probes with minor allele frequency >0.05 (McCartney et al. 2016) were removed for analysis. The final number of CpGs for analysis was 790,026. Quality checks, data normalization, statistical preprocessing, and beta-value calculation, which ranges from 0 to 1 and represents the proportion of unconverted cytosines (Cs) in bisulfite-converted DNA at specific locations, were performed using the R package minfi (Fortin et al. 2017). We estimated Houseman cell proportions (CD8T, CD4T, NK, B cells, monocytes, and granulocytes), also using the R package minfi, to use them as adjustment variables in the regression models. We detected and corrected for potential batch effects by sample plate, sample row, and DNA isolation time with the combat function (*sva* R package). We conducted annotation of CpGs to the nearest gene according to the Illumina Infinium MethylationEPIC Manifest File (version 1.0 B4) (Fortin et al. 2017; Illumina Inc. 2019).

Statistical Methods

DMPs. Four predictors of DMPs were assessed in separate analyses: total urinary Cd concentrations (log-transformed in μg/g of creatinine units), current smoking status (vs. never smoking), former smoking status (vs. never smoking), and cumulative smoking dose (pack-year units). Beta values (i.e., DNAm proportions at a given CpG) were converted into *M* values (logit transformation of DNAm proportions) before conducting linear regression models as implemented with the R package limma (version v3.28.14). Empirical Bayes shrinkage (conducted by limma) has been reported to work better in *M* values than in beta values due to the heteroscedastic nature of the beta values (Teschendorff et al. 2018). In addition, the logistic transformation monotonically maps values in the interval (0,1) to the whole real line $(-\infty,\infty)$ and has been previously used to transform proportion data to fulfill linear modeling assumptions (Chen et al. 2017, Warton and Hui et al. 2011). Thus, we decided to logit-transform β values and then apply linear regression. Models adjusted for biologically relevant variables: age at baseline, sex, study region, BMI, Houseman blood cell composition, and five principal components to account for genetic population stratification (Price et al. 2006). We further adjusted for smoking status (never, former, current) and eGFR in Cd models. We accounted for multiple comparisons by controlling for an FDR (Riancho et al. 2016) at a 0.05 cutoff. Genomic inflation can lead to inflated *p*-values and false positives in EWAS studies (van Iterson et al. 2017), and it was addressed by calculating a set of surrogate variables (SVs) (Leek et al. 2012) from fully adjusted Cd and smoking models using the R package SmartSVA (version 0.1.3) to estimate SVs out of the unexplained DNAm variability and further adjusting the models for them. Although in subsequent bioinformatic analysis we focused on DMPs at an FDR-corrected threshold of statistical significance of 0.05, in the supplementary material we provide the full EWAS output, including nonstatistically significant DMPs (Excel Tables S1–S4).

Mediation analysis. The main question addressed by causal mediation analysis is to what extent is the total effect of changing an exposure *E* mediated by an intermediate variable *M* on the causal pathway to outcome *Y* (the indirect effect) and to what extent is the effect not mediated by *M* (the direct effect) (Pearl 2012; Valeri and Vanderweele 2013). In our mediation framework the "exposure" *E* of interest is smoking, a surrogate for many toxic compounds in tobacco. We know that smoking is an important predictor of DNAm (outcome *Y*), and being a current or former smoker as well as accumulated smoking are related to changes in site-specific DNAm, independently of age, sex, and confounders, in comparison with never smokers. The intermediate variable *M* that we are specifically interested in investigating is urinary Cd, a well-established exposure biomarker, because smoking can cause changes in Cd biomarkers through direct intake through the lungs. Alternatively, the association of Cd biomarker to DNAm is substantiated based on published experimental studies reporting that increased exposure to Cd induces differences in DNAm.

We used the "product of coefficients" method (Pearl 2012) to assess the contribution of increase in urinary Cd concentrations from smoking (current, former, and cumulative smoking in separate models) to the corresponding absolute change in DNAm proportions at specific CpGs (i.e., smoking-related changes in DNAm proportions mediated by Cd). First, for the outcome models, because collapsibility is required for mediation analysis (Valeri and Vanderweele 2013), we used beta regression with an identity link for additively modelling untransformed methylation proportions (instead of the *M*-values) in a generalized linear regression setting with the mgcv package (version 1.8-31) of the R software. Specifically, we fitted beta regression models for DNAm (dependent variable in separate models for each CpG that was significant at 0.05 FDR-corrected *p*-values for Cd and current smoking (excluding former smokers), for Cd and former smoking (excluding current smokers), and for Cd and pack-years, by smoking variables and log-transformed urine Cd (independent variables) and confounders (age, sex, BMI, eGFR, study center, cell counts, and genetic principal components). Second, for the mediator model, we fitted linear regression models for log-transformed urine Cd concentration (dependent variable) by smoking variables (i.e., current and former smoking and pack-years as independent variables in separate models) and the same confounders included in the outcome model. Third, we calculated the effect of smoking on DNAm mediated through Cd (i.e., "indirect" effect) as the product of the mean change in log-Cd concentrations by smoking (smoking regression coefficient from the mediator model) and the absolute change in methylation percentages associated with log-transformed Cd concentrations (Cd regression coefficient from the outcome model). Mediated effects were expressed as the absolute amount of the smoking-associated change in DNAm attributable to Cd ("absolute mediated effect"), as well as the percentage of the smoking-associated change in DNAm attributable to Cd out of the "total effect," defined as the sum of the indirect and direct effects ("relative mediated effect"). Additionally, 95% confidence intervals were derived by simulation from the estimated model coefficients and covariance matrices (adapted from Lange and Hansen 2011).

Sensitivity analyses. We conducted a number of sensitivity analyses. First, we repeated the models for the top 25 significant DMPs for current vs. never smokers in analyses stratified by center and sex. Second, we repeated the DMP analysis for the six significant DMPs for urinary Cd restricting the analysis to never smokers, to assess whether the association between Cd and DNAm was only reflecting smoking status. Third, as we do not have cotinine measurements in this study and smoking is a self-reported variable, we

used the EpiSmokEr tool (Bollepalli et al. 2019) to predict smoking status based on DNAm data. Fourth, we reran the Cd models excluding the eGFR variable as well as additionally adjusting for specific gravity. Last, because probe type bias might be an issue, we applied Regression on Correlated Probes (RCP) normalization after snoob normalization as implemented by the R package ENmix (version 1.8.0) (Niu et al. 2016) to see whether the results changed.

Enrichment analyses. Among Cd- and smoking-related DMPs with a FDR-corrected p-value <0.05, we conducted a hypergeometric analysis to identify enriched gene sets using Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000; Kanehisa et al. 2019) and Gene Ontology (GO) (Ashburner et al. 2000; The Gene Ontology Consortium 2018) as conducted by the gometh function of the Bioconductor missMethyl package (version 1.6.2) (Phipson et al. 2015). Among DMPs with a nominal p-value $<10\times10^{-5}$, we additionally investigated enrichment for molecular regulatory elements of statistically significant DMPs in comparison with nonstatistically significant DMPs, including annotated functional gene regions (TSS1500 and TSS200, 5′UTR, 1 st Exon, Gene body, and 3′UTR) and regions defined by the CpG density context (Open Sea, N-shelf, N-shore, Island, S-shore, and S-shelf). Similarly, we also investigated enrichment for 15-chromatin states retrieved from the ROADMAP Epigenomics Mapping Consortium (ChromHMM version 1.10) (Bernstein et al. 2010; Ernst and Kellis 2017; ROADMAP Epigenomics Project). The mapping of DNAm changes to the most probable chromatin state (combination of histone marks associated to varying degrees of expression) using ChromHMM is a way to identify DNA regulatory activity in a tissue-specific manner, including coding and noncoding portions of the genome, which provides potential functional mechanisms to EWAS findings. For regulatory elements enrichment analysis, p-values were corrected using the Bonferroni method considering 15 chromatin states, 6 categories relatives to functional gene region, and 6 categories relative to CpG density context.

Molecular characterization and* in silico *protein-network analysis of DNAm metabolism. To conduct the molecular characterization of DNAm metabolism we searched KEGG (Kanehisa and Goto 2000) public database for key pathways directly or indirectly related to DNAm (including factors related to methyl groups' depletion) as of 26 June 2019, using the following descriptors: "Glutathione metabolism" (hsa00480), "Citrate cycle" (hsa00020), "One carbon pool by folate" (hsa00670), "Selenocompound metabolism" (hsa00450), and "Cysteine and methionine metabolism" (hsa00270). The search strategy retrieved a total of 158 proteins. TET1, TET2, TET3, SELENOP, AICDA, TDG, and APEX1 genes were not retrieved by our search strategy, but we manually added them to the final list of key proteins because of their potential role in active DNA demethylation according to previous reports (Chia et al. 2011; Lee et al. 2009; Ruiz-Hernandez et al. 2015).

Subsequently, among DMPs with a FDR-corrected p-value <0.05, we created a list of unique protein-coding genes attributed to Cd-DMPs and to the union of the identified smoking related-

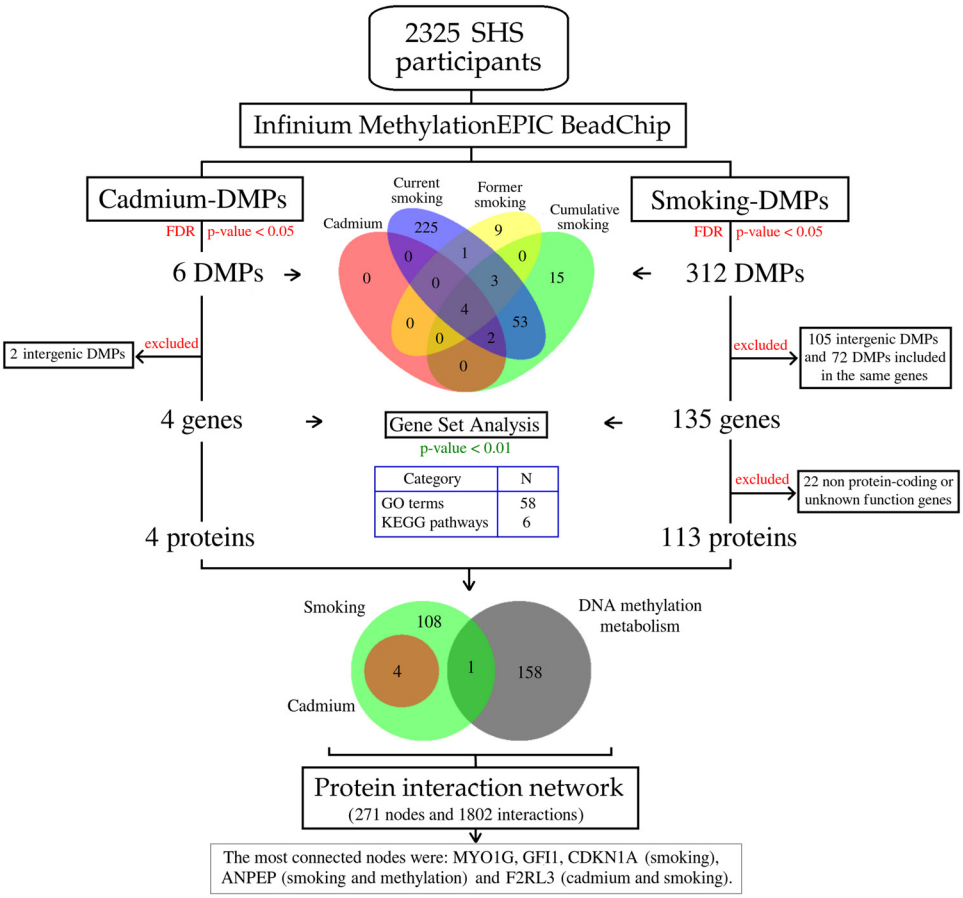

Figure 1. Summary of results from the epigenome-wide association study of cadmium (Cd) and smoking with DNA methylation (DNAm) and subsequent bioinformatics analysis. After considering a false discovery rate (FDR)-corrected p-value < 0.05, we obtained 6 Cd-differentially methylated positions (DMPs) and 312 smoking DMPs in a genomic exploration of ~790,000 CpG sites with the Infinium MethylationEPIC BeadChip. We excluded nonprotein-coding genomic regions and duplicate genes and conducted gene set enrichment analysis. We subsequently evaluated protein interaction networks among protein-encoding genes from the STRING database, resulting in a protein interaction network of 271 nodes and 1,802 interactions. We observed highly connected nodes, which were directly or indirectly related to DNAm metabolism key proteins.

Table 1. Median (IQR) of urine cadmium (μg/g) and accumulated smoking dose (pack-years) levels and percentage of former and current smoking status by participant's characteristics.

	N	Cadmium (μg/g)	Former smoking (%)	Current smoking (%)	Cumulative smoking (pack-years)
Overall	2,325	0.97 (0.62, 1.50)	32.2	38.4	4 (0, 18)
Age (y)					
<50	669	0.89 (0.56, 1.31)	47.3	56.9	4 (0, 17)
50–64	1,246	1.00 (0.64, 1.55)	32.6	30.7	3 (0, 18)
≥65	410	1.03 (0.64, 1.59)	20.1	12.4	2 (0, 20)
Sex					
Men	964	0.71 (0.47, 1.10)	51.1	53.4	9 (0, 27)
Women	1,361	1.16 (0.77, 1.78)	48.9	46.6	1 (0, 12)
Smoking status					
Never	684	0.86 (0.54, 1.36)	0.0	0.0	0 (0, 0)
Former	748	0.81 (0.55, 1.25)	100.0	0.0	7 (2, 20)
Current	893	1.17 (0.77, 1.81)	0.0	100.0	14 (5, 30)
Center					
Arizona	312	0.76 (0.53, 1.19)	15.5	6.3	0 (0, 5)
Oklahoma	981	0.86 (0.55, 1.33)	45.5	37.3	2 (0, 16)
Dakota	1,032	1.11 (0.75, 1.78)	39.0	56.4	7 (0, 23)
Obesity					
No	1,238	1.05 (0.66, 1.63)	44.8	62	6 (0, 23)
Yes	1,087	0.87 (0.57, 1.35)	55.2	38	2 (0, 13)
eGFR (ml/min/1.73 m^2)					
<60	76	0.91 (0.53, 1.51)	32.89	32.89	7 (2, 25)
≥60	2,249	0.97 (0.62, 1.50)	32.15	38.59	11 (3, 26)

Note: IQR, interquartile range.

DMPs (Figure 1). Finally, we constructed a protein interaction network of proteins encoded by the identified smoking and Cd-related genes with respect to the DNAm key proteins selected from the search strategy described above by obtaining protein interactions information from STRING database version 11.0 (Szklarczyk et al. 2019). The STRING database provides a confidence score (from 0 to 1) to indicate the estimated likelihood that the annotated interaction between a given pair of proteins is biologically meaningful, specific, and reproducible, according to the evidence derived from in-house predictions, homology transfers, and the externally maintained databases (Szklarczyk et al. 2019). We displayed protein interaction networks with Cytoscape (version 3.7.1) (Shannon et al. 2003) using the yfiles Organic layout. As a result, 4 and 113 Cd and smoking-related nodes, respectively, were included in the network, in addition to the DNAm related-nodes, which included 159 nodes with high confidence scores (CHAC1, CHAC2, GSTT1, GSTT2, INMT and NAT8B nodes were excluded from the original list of 165 DNAm metabolism selected proteins because they showed connections with low confidence scores).

Results

Descriptive Analysis

The prevalence of current and former smokers was 38.4% and 32.2%, respectively. Median interquartile range (IQR) accumulated smoking and urinary Cd levels were, respectively, 4 (0, 18) cigarette pack-years and 0.97 (0.62, 1.50) μg/g. Older participants had higher levels of Cd, as did women in comparison with men (Table 1). As expected, current smokers showed higher urine Cd concentrations. The difference in Cd levels between smoking status groups was statistically significant (p-value $<2\times10^{-16}$, data not shown). The geometric mean ratio (GMR) of urine Cd comparing current and former smokers with never smokers was 1.45 [95% confidence interval (CI): 1.39, 1.52] and 1.10 (95% CI: 1.04, 1.17), respectively, in models adjusting for age, sex, BMI, center, and eGFR (data not shown). The corresponding GMR per a 10-pack–year increase was 1.10 (95% CI: 1.06, 1.14) (data not shown).

Differentially Methylated Positions

We found differential methylation in 6 CpGs for urinary Cd at the FDR-corrected significance level of 0.05 (Table 2; Figure S2; Excel Table S1). Two CpGs (cg14391737 and cg17739917) were new from the Illumina Infinium MethylationEPIC array and therefore have not been reported before. One CpG was annotated to *AHRR*, one to *F2RL3*, one to *PRSS23*, two to *2q37.1*, and one to *RARA*. Among those six significant CpGs in the overall population, four remained significant after restricting the models to never smokers (Table 2). A total of 288 CpGs were differentially methylated in relation to current smoking (Table 3; Figure S2; Excel Table S2). A total of 149 CpGs were new from the Illumina Infinium MethylationEPIC array. The top DMPs annotated to smoking-related genes in a previous meta-analysis was replicated in our study (Table 3, Joehanes et al. 2016). Among the top 25 DMPs for current smoking (Table 3), 5 were annotated to *AHRR*, 2 to *F2RL3*, 1 to *PRSS23*, 3 to *2q37.1*, and 1 to *RARA*. A total of 17 CpGs were differentially methylated in relation to former smoking (8 of them were also associated with current smoking) (Figure S2; Table S2; Excel Table S3), 5 of them being new from the Illumina Infinium MethylationEPIC array. A total of 77 CpGs were differentially methylated in relation to cigarette pack-years [62 of them also associated with current smoking (Figure S2; Table S3; Excel Table S4), and 29 of them being new from the Illumina Infinium MethylationEPIC array]. The genomic inflation factors (GIF) associated with the nominal p-values for the Cd, current and former smoking, and pack-years analyses before SV adjustment were 1.49, 2.91, 1.02, and 1.30, respectively. In the final models, which included SVs adjustment, the corresponding GIFs were, respectively, 1.05, 1.02, 1.02, and 1.003.

Sensitivity Analyses

Results were similar when excluding the eGFR variable from the Cd models. However, we still wanted to keep eGFR in the final model because the filtration rate affects metal levels excreted in the urine (Zheng et al. 2015). Adjusting for specific gravity slightly attenuated the results but did not change them in a substantial way.

Table 2. Top differentially methylated positions (DMPs) associated to log-transformed urine cadmium (Cd) levels (i.e., Cd DMPs).

					Overall			Never smokers		
CpG	Chr	Position	Gene	Function	Odds ratio (95% CI)[a]	Nominal p-value	FDR p-value	Odds ratio (95% CI)[a]	Nominal p-value	FDR p-value
cg14391737[b]	chr11	86513429	PRSS23	Encodes Serine protease 23. May be an important ovarian protease	0.92 (0.9, 0.94)	1.01×10^{-12}	6.18×10^{-7}	0.94 (0.89, 0.98)	0.003	0.006
cg21566642	chr2	233284661	2q37.1	Uncharacterized	0.92 (0.9, 0.94)	1.56×10^{-12}	6.18×10^{-7}	0.97 (0.92, 1.01)	0.1	0.2
cg05575921	chr5	373378	AHRR	Mediates dioxin toxicity. Involved in cell growth and differentiation	0.84 (0.8, 0.88)	1.12×10^{-11}	2.96×10^{-6}	0.85 (0.77, 0.94)	0.002	0.006
cg01940273	chr2	233284934	2q37.1	Uncharacterized	0.96 (0.94, 0.97)	2.41×10^{-8}	0.004	0.98 (0.95, 1.02)	0.3	0.3
cg03636183	chr19	17000585	F2RL3	Blood coagulation, inflammation, and response to pain	0.96 (0.94, 0.97)	1.60×10^{-7}	0.02	0.96 (0.93, 1.00)	0.03	0.04
cg17739917[b]	chr17	38477572	RARA	Development, differentiation, apoptosis, granulopoeisis, clock genes transcription	0.95 (0.93, 0.97)	2.63×10^{-7}	0.03	0.93 (0.88, 0.97)	0.002	0.006

Note: BMI, body mass index; CI, confidence interval; FDR, false discovery rate.
[a]Linear regression models were fitted using logit-transformed DNAm proportions as dependent variables separately for each CpG and were adjusted for smoking status (never, former, current), age (years), sex (male/female), BMI (kg/m^2), Houseman cell proportions (CD8T, CD4T, NK, B cells, monocytes and granulocytes), study center (Arizona, Oklahoma or North and South Dakota), five genetic principal components and estimated glomerular filtration rate (ml/min/1.73 m^2).
[b]Not available CpGs in the Illumina HumanMethylation450 BeadChip array.

In stratified analysis by sex and center for the top 25 DMPs in current smoking models, all the DMPs remained statistically significant and directionally consistent (Tables S4 and S5) except for those from Arizona, where the results where attenuated and 4 of the 25 sites did not reach the statistical significance, probably due to the much smaller sample size in comparison with Oklahoma and North and South Dakota. The contingency table of the self-reported smoking status vs. predicted values by EpiSmokEr can be found in Table S8. Overall, only 1,179 of the 2,325 patients were classified correctly. However, most of the people predicted to be current smokers also reported to be current smokers. We saw similar results when applying RCP and snoob normalizations to those obtained when only applying snoob normalization (data not shown).

Mediation Analysis

The number of significant DMPs associated with smoking decreased from 288 to 247 when adjusting for urinary Cd levels (data not shown). The mediated effect in the relationship between current smoking and DNAm through urinary Cd was statistically significant for the six CpGs that were commonly differentially methylated for both Cd and current smoking using the FDR-corrected p-values (Table 4). Median (minimum, maximum) total effects for the association between current smoking and DNAm were -6.83% (-13.19%, -5.70%). The corresponding absolute mediated effects through Cd were -0.70% (-0.88%, -0.44%), respectively. Total effects for former smoking (four CpGs) and cigarette pack-years (six CpGs) and corresponding mediated effects by Cd are shown in Tables S6 and S7. Excel Tables S5, S6, and S7 show mediation analysis for each of the CpGs associated with the smoking variables (current smoking, former smoking, and pack-years).

Enrichment Analyses

Among smoking- and Cd-DMPs with an FDR p-value <0.05 (6 DMPs for Cd; 288 and 17, respectively, for current and former smoking status; and 77 for cigarette pack-years), we identified a total of 58 GO and 6 KEGG enriched pathways (out of 17,479 GO and 336 and KEGG available gene sets), respectively. The full list of evaluated GO terms and KEGG pathways can be found in Excel Tables S8 and S9. The top 10 GO terms and KEGG pathways were related to cancer (GO:0007229; GO:0071498; GO:0010595; GO:0060574; GO:0005886; GO:0060534; GO:0043149; GO:0002068; hsa05206; hsa04072; hsa05200; hsa05202), CVD and risk factors (GO:0071498; GO:0010595; GO:0060534; GO:0007155; GO:0043149; GO:0002068; GO:0072678; hsa05202; hsa04070; hsa04810), inflammation (GO:0071498; GO:0010595; GO:0007155; GO:0043149); and others (Tables S8 and S9).

Among DMPs with a nominal p-value $<10^{-5}$ (40,443 for Cd; 44,092 for current smoking; 42,000 for former smoking; and 45,612 for pack-years), the molecular enrichment relative to CpG density context (Figure S3, Panel A) showed underrepresentation of CpG Islands for current smoking [odds ratio (OR) $=0.36$, $p=2.4\times10^{-5}$] and open sea (OR $=0.26$, $p=5.3\times10^{-5}$) for former smoking DMPs. However, CpGs located in S. Shores showed an overrepresentation in former smoking-DMPs (OR $=2.96$, $p=7.4\times10^{-3}$) (Figure S3, Panel B). The enrichment analysis relative to functional gene region revealed some significant findings (Figure S3, Panel B). For current smoking-DMPs, we observed an overrepresentation of sites located at 5′UTR (OR $=1.60$, $p=0.02$) and underrepresentation at first exon (OR $=0.50$, $p=0.02$), TSS200 (OR $=0.08$, $p=7.1\times10^{-8}$) and TSS1500 (OR $=0.57$, $p=0.03$). Cumulative smoking-DMPs showed a statistically significant underrepresentation of TSS200 (OR $=0.10$, $p=5.6\times10^{-3}$). We observed a strong enrichment of several chromatin states for all smoking-related analyses (Figure 2). In particular, there was a consistent over-representation of CpGs in enhancer for current (OR $=4.76$, $p=4.8\times10^{-9}$), former (OR $=2.88$, $p=5.9\times10^{-3}$) and cumulative (OR $=4.81$, $p=4.9\times10^{-7}$) smoking-DMPs.

In Silico *Protein-Interaction Analysis*

The list of candidate proteins for DNAm-related pathways are shown in Excel Table S10. From the protein–protein interaction, we obtained a network of 271 nodes with 1,802 interactions (Figures 1 and 3). Among nodes with more than 20 interactions, we mostly found smoking-only nodes (*MYO1G, GFI1*, *CDKN1A*, *GNG12*, and *CYP1B1*), smoking and methylation metabolism nodes (*ANPEP*), and 1 Cd and smoking node (*F2RL3*). Seven

Table 3. Top 25 differentially methylated positions (DMPs) comparing current smoking status with never smoking status (i.e., current smoking-DMPs).

CpG	Chr	Position	Gene	Function	Odds ratio (95% CI)[a]	Nominal p-value	FDR p-value	FDR p-value in meta-analysis (Joehanes et al. 2016)[b]	Direction of association in meta-analysis (Joehanes et al. 2016)[b]
cg05575921	chr5	373378	AHRR	Mediates dioxin toxicity. Involved in cell growth and differentiation	0.38 (0.35, 0.42)	1.99×10^{-99}	1.57×10^{-93}	6.1×10^{-22}	Inverse
cg21566642	chr2	233284661	2q37.1	Uncharacterized	0.71 (0.69, 0.74)	1.92×10^{-67}	7.58×10^{-62}	4.5×10^{-21}	Inverse
cg01940273	chr2	233284934	2q37.1	Uncharacterized	0.8 (0.78, 0.83)	1.37×10^{-58}	3.60×10^{-53}	9.8×10^{-30}	Inverse
cg17739917	chr17	38477572	RARA	Development, differentiation, apoptosis, granulopoeisis, clock genes transcription	0.75 (0.73, 0.78)	1.90×10^{-56}	3.75×10^{-51}	Not available	—
cg03636183	chr19	17000585	F2RL3	Blood coagulation, inflammation, and response to pain	0.8 (0.78, 0.83)	5.84×10^{-52}	9.23×10^{-47}	5.7×10^{-17}	Inverse
cg26703534	chr5	377358	AHRR	Mediates dioxin toxicity. Involved in cell growth and differentiation	0.83 (0.81, 0.86)	1.84×10^{-46}	2.42×10^{-41}	7.2×10^{-18}	Inverse
cg21161138	chr5	399360	AHRR	Mediates dioxin toxicity. Involved in cell growth and differentiation	0.86 (0.84, 0.88)	3.82×10^{-28}	4.31×10^{-23}	7.9×10^{-13}	Inverse
cg25648203	chr5	395444	AHRR	Mediates dioxin toxicity. Involved in cell growth and differentiation	0.84 (0.81, 0.87)	4.92×10^{-25}	4.86×10^{-20}	2.7×10^{-11}	Inverse
cg09935388	chr1	92947588	GFI1	Transcriptional repressor. Plays a role in hematopoiesis and oncogenesis	0.75 (0.71, 0.79)	6.17×10^{-25}	5.42×10^{-20}	7.0×10^{-14}	Inverse
cg14391737	chr11	86513429	PRSS23	Encodes Serine protease 23. May be an important ovarian protease	0.81 (0.78, 0.84)	2.18×10^{-24}	1.72×10^{-19}	Not available	—
cg19859270	chr3	98251294	GPR15	Chemokine receptor for human immunodeficiency virus type 1 and 2	0.84 (0.81, 0.87)	2.99×10^{-24}	2.15×10^{-19}	6.3×10^{-17}	Inverse
cg17287155	chr5	393347	AHRR	Mediates dioxin toxicity. Involved in cell growth and differentiation	0.78 (0.74, 0.82)	2.16×10^{-23}	1.42×10^{-18}	1.1×10^{-13}	Inverse
cg14753356	chr6	30720108	Intergenic	Uncharacterized	0.86 (0.83, 0.89)	4.34×10^{-21}	2.64×10^{-16}	2.3×10^{-14}	Inverse
cg16841366	chr2	233286192	2q37.1	Uncharacterized	0.83 (0.8, 0.87)	5.54×10^{-21}	3.13×10^{-16}	Not available	—
cg18110140	chr15	75350380	Intergenic	Uncharacterized	0.87 (0.84, 0.9)	6.71×10^{-20}	3.53×10^{-15}	Not available	—
cg21911711	chr19	16998668	F2RL3	Blood coagulation, inflammation, and response to pain	0.85 (0.82, 0.88)	1.26×10^{-19}	6.22×10^{-15}	Not available	—
cg25189904	chr1	68299493	GNG12	Modulator in transmembrane signaling systems	0.81 (0.78, 0.85)	2.13×10^{-19}	9.90×10^{-15}	3.3×10^{-18}	Inverse
cg22675726	chr18	3179889	MYOM1	Structural constituent of muscle. Associated with hypertrophic cardiomyopathy	0.85 (0.82, 0.88)	4.70×10^{-19}	2.06×10^{-14}	Not available	—
cg00045592	chr1	160714299	SLAMF7	Modulator on the activation and differentiation of immune cells	0.89 (0.86, 0.91)	2.37×10^{-18}	9.87×10^{-14}	Not available	—
cg00475490	chr11	86517110	PRSS23	Encodes Serine protease 23. May be an important ovarian protease	0.78 (0.74, 0.83)	6.60×10^{-18}	2.61×10^{-13}	Not available	—
cg16522223	chr22	43400443	PACSIN2	Intracellular vesicle-mediated transport	0.93 (0.91, 0.94)	3.23×10^{-17}	1.22×10^{-12}	Not available	—
cg09338374	chr22	39888390	Intergenic	Uncharacterized	1.1 (1.07, 1.12)	6.59×10^{-17}	2.37×10^{-12}	Not available	—
cg15159987	chr19	17003890	CPAMD8	Innate and acquired immunity	0.93 (0.91, 0.95)	1.50×10^{-16}	5.16×10^{-12}	4.6×10^{-5}	Inverse
cg21322436	chr7	145812842	CNTNAP2	Cell adhesion. Related to autism	0.87 (0.84, 0.9)	1.59×10^{-16}	5.25×10^{-12}	9.7×10^{-15}	Inverse
cg26768182	chr9	134272679	Intergenic	Uncharacterized	0.88 (0.85, 0.9)	1.72×10^{-16}	5.44×10^{-12}	Not available	—

Note: —, no data; BMI, body mass index; CI, confidence interval; DNAm, DNA methylation; FDR, false discovery rate.

[a]Linear regression models were fitted using logit-transformed DNAm proportions as dependent variables separately for each CpG and were adjusted, age (y), sex (male/female), BMI (kg/m^2), Houseman cell proportions (CD8T, CD4T, NK, B cells, monocytes and granulocytes), five genetic principal components and study center (Arizona, Oklahoma, or North and South Dakota).

[b]Not available CpGs in the metanalysis by (Joehanes et al. 2016) reflect that the metanalysis was based in the Illumina Infinium HumanMethylation450 BeadChip array.

Table 4. Changes in DNA methylation percentages comparing current with never smokers attributable to changes in urinary Cd concentrations ("mediated effect").

					Outcome model[a]		Mediated (i.e., indirect) effects[b] of current vs. never smoking through Cd from smoking		Total effects of current vs. never smoking in DNA methylation %
CpG	Chr	Position	Gene	Function	Absolute change in DNAm % per log-Cd unit increase (95% CI)	Absolute change in DNAm % comparing current vs. never smokers ("Direct effect") (95% CI)	Difference in change attributable to Cd (95% CI)	Percentage of change attributable to Cd (95% CI)	Sum of direct and indirect effects (95% CI)
cg14391737	chr11	86513429	PRSS23	Encodes Serine protease 23. May be an important ovarian protease	−2.32 (−2.87, −1.78)	−5.84 (−6.57, −5.1)	−0.88 (−1.14, −0.63)	13.05 (9.4, 17.18)	−6.72 (−7.44, −5.99)
cg05575921	chr5	373378	AHRR	Mediates dioxin toxicity. Involved in cell growth and differentiation	−2.24 (−2.48, −2.01)	−12.34 (−12.78, −11.89)	−0.85 (−1.02, −0.69)	6.43 (5.23, 7.72)	−13.18 (−13.64, −12.73)
cg21566642	chr2	233284661	2q37.1	Uncharacterized	−1.97 (−2.52, −1.42)	−8.72 (−9.47, −7.97)	−0.74 (−1, −0.51)	7.82 (5.4, 10.6)	−9.46 (−10.2, −8.73)
cg03636183	chr19	17000585	F2RL3	Blood coagulation, inflammation and response to pain	−1.73 (−2.17, −1.29)	−5.04 (−5.64, −4.45)	−0.65 (−0.87, −0.46)	11.45 (8.12, 15.33)	−5.7 (−6.28, −5.11)
cg17739917	chr17	38477572	RARA	Development, differentiation, apoptosis, granulopoeisis, clock genes transcription	−1.49 (−2.03, −0.95)	−6.39 (−7.13, −5.66)	−0.56 (−0.8, −0.35)	8.06 (4.97, 11.62)	−6.95 (−7.67, −6.24)
cg01940273	chr2	233284934	2q37.1	Uncharacterized	−1.16 (−1.55, −0.77)	−5.29 (−5.82, −4.76)	−0.44 (−0.61, −0.28)	7.61 (4.86, 10.73)	−5.73 (−6.24, −5.21)

Note: BMI, body mass index; Cd, cadmium; CI, confidence interval; DNAm, DNA methylation.

[a]Absolute changes (mean differences) in methylation percent comparing current with never smokers (i.e., "Direct effect") and absolute changes in DNAm % associated with one-unit increase in log-transformed urine Cd were both estimated as the regression coefficients from beta models with an identity link (i.e., "Outcome model") which included the following independent variables: current smoking status ("Exposure"), log-transformed Cd ("Mediator"), age (years), sex (men/women), BMI, Houseman cell proportions (CD8T, CD4T, NK, B cells, monocytes and granulocytes), five genetic principal components, and study center (Arizona, Oklahoma, or North and South Dakota).

[b]Effects mediated through Cd (i.e., "Indirect effects") were calculated using the "product of coefficients" method that multiplies the coefficient for the mean difference in log-transformed cadmium concentrations comparing current with never smokers (i.e., 0.38 [95% CI: 0.32, 0.44] corresponding to coefficient [95% CI] for current smoking status from the linear regression "mediator model" where the outcome is log-transformed Cd adjusted for the same variables as the outcome model, data not shown) by the absolute change in DNAm % associated with one-unit increase in log-Cd concentrations indirect effects were expressed in absolute terms (difference in smoking-related DNAm % change) and relative to the "Total effect" [DNAm % changes (95% CIs) comparing current with never smokers] calculated as the sum of the direct and indirect effects. Note that if the mediation analysis assumptions hold, the estimated "Total effects" in the setting of the "product of coefficients" method is numerically identical to the regression coefficient associated with current smoker status in a beta regression as in the outcome model without Cd.

To illustrate the estimation of mediated effects in the setting of the product of coefficients methods, we provide an example: In absolute terms, the mediated effect of Cd in cg14391737 (column 8) was obtained as $-2.32\,\%$ (column 6) $\times 0.38$ (from the mediator model mentioned above) $= -0.88\,\%$. The corresponding mediated effect in relative terms (column 9) was $-0.88\,\%/-6.72\,\%$("Total effect" from column 10) $\times 100 = 13.1\,\%$. The 95% CIs for the total and mediated effects were derived by simulation from the estimated model coefficients and covariance matrices.

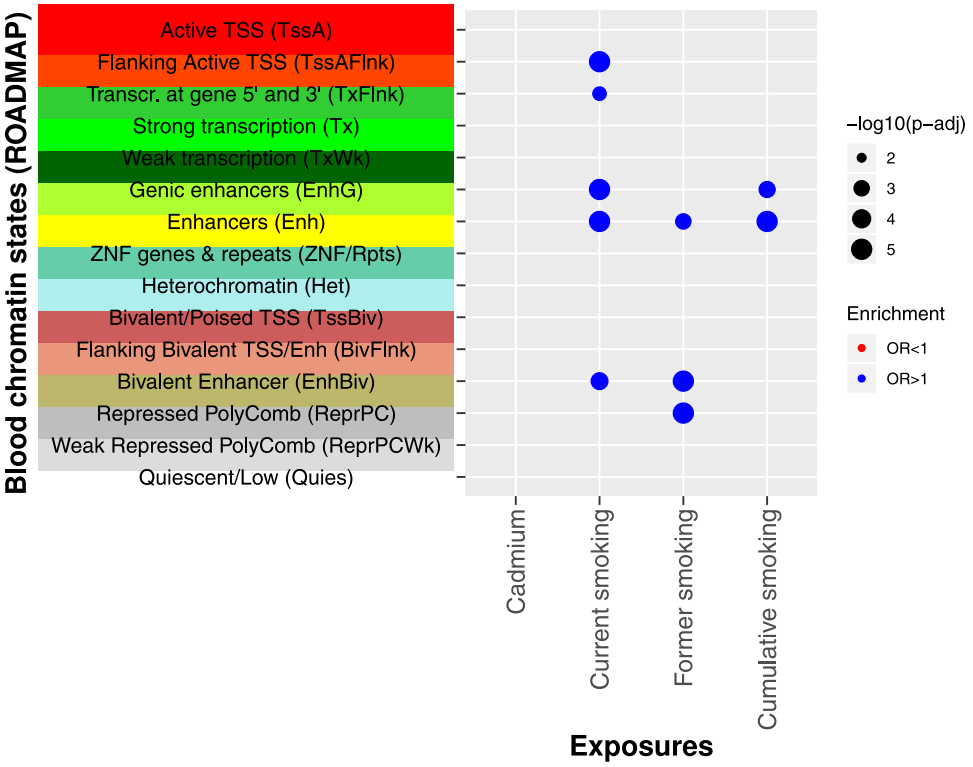

Figure 2. Enrichment analysis of top significant cadmium (Cd) and smoking-related differentially methylated positions (DMPs) (p-value $< 10^{-5}$) for 15-chromatine states from the ROADMAP project. The area of the solid dots is directly proportional to the strength of the statistical evidence in favor of being annotated to a given chromatin state category comparing statistically significant vs. nonstatistically significant DMPs in our data.

smoking-only nodes (*GFI1*, *CYP1B1*, *NFE2L2*, *TSHZ1*, *SIN3B*, *ALPP*, and *ARRB1*) were directly connected to 27 DNAm metabolism nodes. The four Cd and smoking nodes were indirectly connected to DNAm metabolism nodes through six smoking-only nodes. The Cytoscape session is provided so that readers can interactively explore the network (Cytoscape Additional_file.cys supplementary file; publicly available Cytoscape software can be downloaded at https://cytoscape.org/index.html), and the general statistics from the Cytoscape file can be found in Excel Tables S11 (network nodes) and S12 (network edges).

Discussion

After a conservative FDR correction, we found DMPs associated with current smoking, cumulative smoking dose, and urinary Cd levels, being all Cd-DMPs also smoking-DMPs. We replicated the association of well-established smoking-associated DNAm sites, including sites mapped to the genes *AHRR*, *F2RL3*, *PRSS23*, *RARA*, and 2q37.1. In addition, we found several smoking-associated DMPs that had not been previously reported. Our mediation analysis supports that Cd is a partial mediator of the association between smoking and differential DNAm at specific sites. The integrative *in silico* analysis provides further support for interconnected epigenetic mechanisms of both Cd- and smoking-associated health effects, including the possibility of Cd influencing smoking-related DNAm changes at relevant loci.

A large evidence base supports the association between smoking and differential methylation of several genes, with *AHRR*, *F2RL3*, and 2q37.1 being the most documented (Bojesen et al. 2017; Fasanelli et al. 2015; Gao et al. 2015; Joehanes et al. 2016; Reynolds et al. 2015; Zhang et al. 2013b). Little is known about the 2q37.1 region. Although it is one of the most commonly deleted subtelomeric regions (Leroy et al. 2013), the health implications of the differential methylation in this region are not clear. The two Cd-related DMPs located in this region were no longer significant when restricting the Cd models to never smokers, whereas all the other DMPs were still significant among never smokers.

Whether epigenetic modifications are a biological mechanism for well-known smoking-related health effects remains a focus of ongoing investigations. In fact, altered DNAm of *AHRR* and *F2RL3* has been suggested to be part of the biological links between cigarette smoking and several diseases, especially lung cancer and atherosclerosis (Reynolds et al. 2015), in addition to being strong predictors of all-cause mortality in several studies (de Smith et al. 2017; Fasanelli et al. 2015; Zhang et al. 2014). Other smoking-related DMPs in our data such as *CPOX*, *KIAA0087*, *PRSS23*, *GPR15*, and *RARA*, among others, have also been consistently documented (Harlid et al. 2014; Tsaprouni et al. 2014).

Two different Illumina microarrays have been used in previous studies, the Illumina Infinium HumanMethylation27 BeadChip array (which interrogates DNAm at 27,000 CpGs located mostly in promoter regions) and Illumina HumanMethylation450 BeadChip array (including 450,000 CpGs). This is the first human study to use the Illumina Infinium MethylationEPIC BeadChip array to assess differential DNAm in relation to smoking, which led to novel insights. For instance, some new genes that warrant further investigation include a zinc finger protein (ZNF83), with a role of coordinating zinc ions (Cassandri et al. 2017), *PTPN1*, which is involved in oncogenic transformations (Zhu et al. 2007), and *RAB32*, a member of the RAS oncogene family, whose overexpression has been shown in 20% to 25% of all human tumors and up to 90% in pancreatic cancer (Downward 2003). For former smoking, however, we only found six significant CpGs, which indicates that current smoking is more relevant for methylation changes than former smoking. In fact, several studies suggested that methylation changes associated to smoking can be reversed on cessation

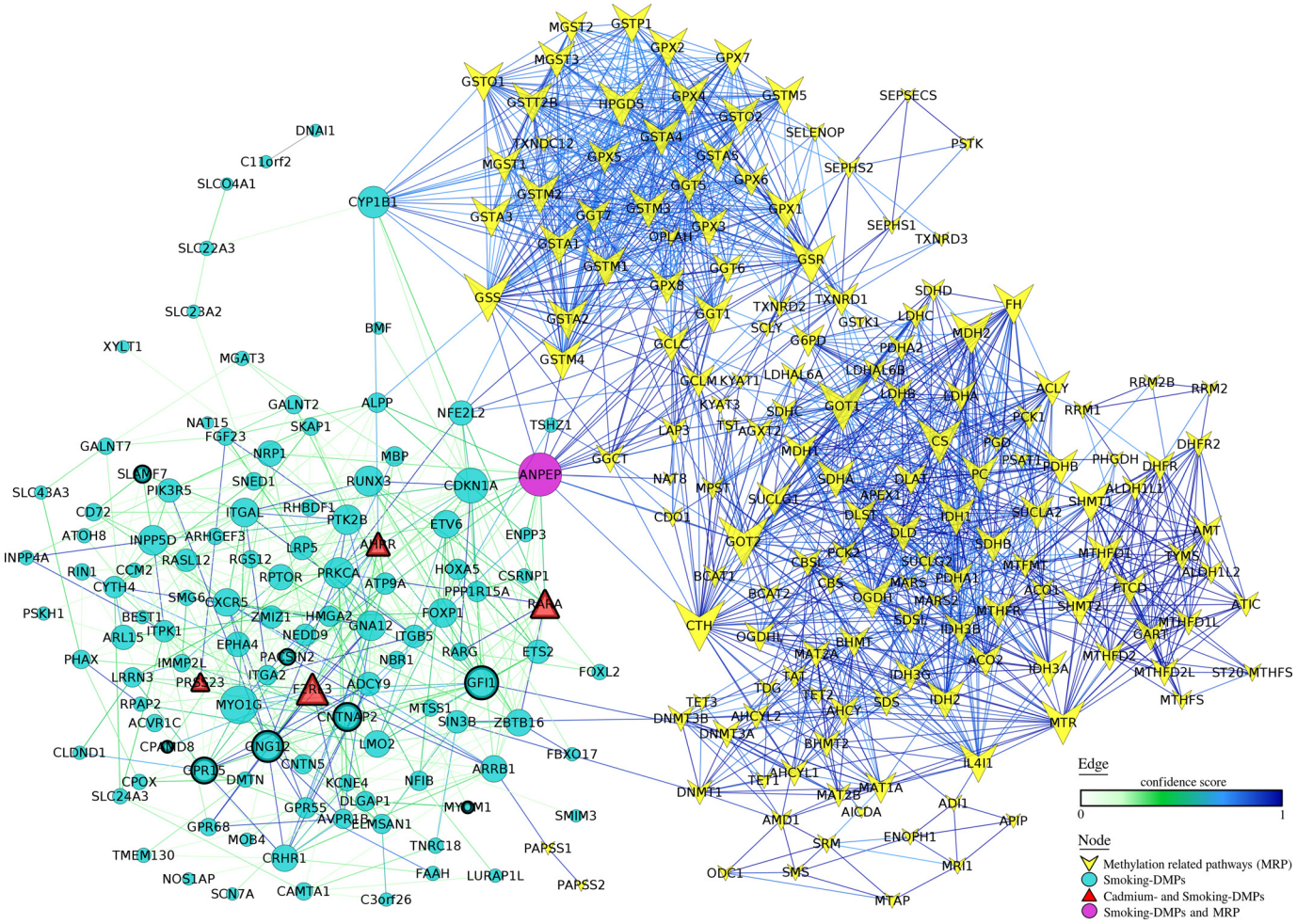

Figure 3. Protein-interaction network between proteins attributed to cadmium (Cd)- and smoking-differentially methylated positions (DMPs) and key proteins in DNA methylation (DNAm) metabolism pathways from KEGG. The nodes correspond to proteins involved in DNAm-related pathways, and proteins encoded by genes associated to Cd-DMPs and smoking-DMPs. The size of the nodes is proportional to the number of connections. The STRING database provides a confidence score (from 0 to 1) to indicate the estimated likelihood that the annotated interaction between a given pair of proteins is biologically meaningful, specific, and reproducible, according to the evidence derived from in-house predictions, homology transfers, and the externally maintained databases (Szklarczyk et al. 2019). Increasingly darker solid edge lines indicate a protein interaction with increasingly higher confidence scores. The nodes included among the top 25 statistically significant Cd- and/or smoking-DMPs in our EWAS show thicker node lines.

(Li et al. 2018; Tsaprouni et al. 2014). It is interesting to note that significant current, former, and cumulative smoking-DMPs were enriched in several categories of blood-specific enhancer regions. These findings support the hypothesis that identified DMPs may be biologically relevant, because it is known that enhancers regulate gene expression in a tissue-specific manner, in this case blood, which is our target tissue. Overall, our findings from molecular regulatory elements enrichment analysis are consistent with the hypothesis of epigenetic dysregulations being involved in mechanisms by which smoking induces diseases.

An interesting aspect is that Cd exposure has been consistently associated with several human health disorders in common with smoking, including CVD (Chowdhury et al. 2018; Tellez-Plaza et al. 2013a) and several types of cancer (García-Esquinas et al. 2014; McElroy et al. 2017), as well as kidney disease (Grau-Perez et al. 2017; Navas-Acien et al. 2009; Orr and Bridges 2017), bone disease (Akesson et al. 2006; Chen et al. 2009), and premature mortality (Larsson and Wolk 2016). Cd has also been related to biomarkers of inflammation (Fagerberg et al. 2017) and oxidative stress (Domingo-Relloso et al. 2019; Nemmiche 2016) in epidemiological studies. Given that smoking is known to be a major source of Cd exposure (Nordberg et al. 2015), several studies have assessed whether Cd could be a mediator between smoking and adverse health effects. In a study conducted in a Swedish cohort (Andersson et al. 2018), about two-thirds of the association between smoking and atherosclerosis was found to be mediated by Cd. In a representative sample of the U.S. general population, Cd partially explained the association of smoking on peripheral arterial disease, which is characterized by the atherosclerotic narrowing of the arteries of the lower extremities (Navas-Acien et al. 2004). Cd also displayed a potential role as mediator of the association between smoking and lung cancer (García-Esquinas et al. 2014).

To our knowledge, however, no studies have evaluated the potential mediating role of Cd in the association between smoking and DNAm. This hypothesis is relevant, given the increasing evidence on the connection of DNAm with relevant biological pathways by which environmental exposures could cause disease. Although some studies have assessed the potential effects of Cd exposure in DNAm, many of them were conducted in experimental settings or interrogated only a reduced number of specific CpGs and genes (Ruiz-Hernandez et al. 2015). Few EWAS, however, have been conducted for Cd exposure in adults. None of our DMPs for Cd have been found in previous studies; nevertheless, these studies were not comparable with ours. One EWAS

involved placental Cd and placental DNAm and is therefore not comparable to blood DNAm given the tissue-specificity of DNAm (Everson et al. 2018). Another EWAS specifically looked at differences in DNAm age by urinary Cd concentrations (Demanelis et al. 2017). A study evaluated DNAm differences associated with Cd in 4.6 million CpGs. However, the sample size was very small ($n = 17$), and none of the sites were significant after correcting for multiple comparisons, so top hits may be unstable (Sanders et al. 2014). In our data, changes in Cd levels from smoking were associated to DNAm changes in the most broadly known smoking-associated genes, suggesting common epigenetic dysregulations due to smoking and Cd exposure at these sites. Nonetheless, only 6 DMPs were found for Cd as opposed to the 288 found for current vs. never smoking. There are more than 7,000 chemicals in cigarette smoke, from which 93 are classified as harmful and potentially harmful constituents (HPHCs), which include Cd and other metals (Morgan et al. 2017). More studies exploring the epigenetic effects of other harmful chemicals from smoking are needed.

The protein–protein interaction network, on the other hand, deepens into the potentially close relationship between DNAm metabolism pathways and genes attributed to smoking and Cd-DMPs. For instance, the Aminopeptidase N (*ANPEP*), which belongs to the glutathione metabolism pathway showed the highest number of connections (directly interacting with 16 other methylation-related proteins and 15 smoking nodes including *F2RL3* and *RARA*, also associated with Cd). It is interesting to note that negative *APN* (the ANPEP encoded protein) immunoreactivity has been suggested as a prognostic factor for prostate cancer (Sørensen et al. 2013), which has been associated with both Cd and smoking (García-Esquinas et al. 2014; U.S. Department of Health and Human Services 2010). Moreover, the *F2RL3* gene, which encodes a protease-activated receptor that plays a role in blood coagulation, inflammation, response to pain, and cancer, is also highly connected in the network and directly interacts with *RARA*, which has been associated with increased sensitivity to retinoids in T-cell lymphoma in experimental settings (Wang et al. 2017). Other highly connected smoking-DMPs in the network, such as *MYO1G*, *GFI1*, *CDKN1A*, *GNG12*, and *CYP1B1*, have been previously reported as smoking-DMPs in other epigenetic studies (Gonseth et al. 2016; Steenaard et al. 2015; Stueve et al. 2017; Wiklund et al. 2019) and have also been implicated in cancer (Groth-Pedersen et al. 2012; Möröy and Khandanpour 2019; Yao et al. 2018) or cardiometabolic diseases (Dempsie et al. 2013; Parmar et al. 2018). All of them are directly connected to Cd-associated nodes.

Our study has some limitations. First, because smoking is a major source of Cd, and smoking status was based on self-report, a possible concern is whether Cd serves as a biomarker of tobacco exposure rather than reflecting Cd per se. However, in a confirmatory sensitivity analysis for the Cd-associated DMPs among never smokers, we found that four of the six signals remained significant after excluding current and former smokers. This supports the hypothesis that Cd in urine could come from other exposure sources such as diet. A previous study (Olmedo et al. 2017) evaluated the association between several dietary products and urinary Cd levels in the SHS, reporting the potential importance of processed meat products as a dietary source of Cd. Occupational exposure to Cd occurs mainly during mining, smelting, and work with Cd-containing minerals (ATSDR 2012). We do not have occupational data in our study; however, some people might do handmade work with minerals and Cd pigments (e.g., jewelry making) and might be exposed to Cd that way. In addition, secondhand tobacco smoke exposure could also be related to urinary Cd levels in never smokers. In a sensitivity analysis with EpiSmokEr, most of the misclassification was from participants who self-reported to be former smokers but for which the algorithm predicted them to be never smokers. Given that it is unlikely that people would indicate they smoked in the past without doing it (i.e., reporting to be a former smoker when he or she is a never smoker), this aspect could be related to the previously reported reversible nature of some of the smoking-related epigenetic marks (Reynolds et al. 2015). In the same way, many self-reported current smokers were predicted to be former smokers by the algorithm. This prediction could be related to the small number of cigarettes per day smoked by many smokers in our population (median cigarette pack was 4, which would mean that one person could have smoked, for example, 2 cigarettes per day through 40 y, or 4 cigarettes per day through 20 y). Finally, we cannot discard misclassification due to secondhand smoke, because in this study there was a median (IQR) of 1 (0, 6) h of self-reported exposure to secondhand smoke per day. In addition, mediation analysis is subject to strong assumptions (Pearl 2012; Valeri and Vanderweele 2013), for instance, the assumption of no unmeasured confounders. Residual confounding cannot be discarded in observational studies. It is important to note that we obtained essentially identical mediated effects when we used the difference of coefficient methods instead of the product of coefficient methods (data not shown), which indicates that the no exposure-mediator interaction assumption holds in our data.

Another limitation is the fact that we conducted mediation analysis in the setting of a cross-sectional study design. However, cigarettes are known to contain Cd, so the directionality between smoking and Cd exposure relation over time is clear. The directionality between Cd exposure and blood DNAm is less clear, but—given the long half-life of urinary Cd (15–30 y) and the fact that Cd is an exogenous nonessential metal—it is likely that Cd exposure precedes blood DNAm. Finally, although replication of results in independent study populations is required in studies conducting genomic explorations of the methylome, this concern may be a relatively minor one in our case, given the robustness of our data with respect to well-replicated smoking-related signals across a wide range of other population studies (Gao et al. 2015; Joubert et al. 2012).

The strengths of our study include the large sample size, including methylation data for the largest amount of CpGs possible with current microarray technology (Illumina Infinium MethylationEPIC BeadChip); the high quality of the study protocols; and the availability of information to account for potential confounders.

In conclusion, this work supports the hypothesis that Cd, one of the principal toxicants in cigarettes, could be partially explaining well-established associations of smoking with DNAm changes. Cd exposure could thus be linked to smoking adverse health outcomes through epigenetic mechanisms. Although smoking exposure has declined in recent decades, public health interventions are still needed for controlling smoking exposure among the population. Additional strategies are also needed to reduce Cd exposure from other sources beyond tobacco (Ruiz-Hernandez et al. 2017). Findings from our *in silico* study needs functional confirmation in experimental settings, as relevant proteins in our protein-interaction network can contribute to the understanding of Cd and smoking-related adverse health outcomes and mechanisms.

Acknowledgments

This work was supported by grants by the National Heart, Lung, and Blood Institute (NHLBI) (under contract numbers 75N92019D00027, 75N92019D00028, 75N92019D00029, & 75N92019D00030) and previous grants (R01HL090863, R01HL109315, R01HL109301, R01HL109284, R01HL109282,

and R01HL109319 and cooperative agreements U01HL41642, U01HL41652, U01HL41654, U01HL65520, and U01HL65521), by the National Institute of Health Sciences (R01ES021367, R01ES025216, P42ES010349, P30ES009089), by the Spanish Funds for Research In Health Sciences, Carlos III Health Institute, co-funded by European Regional Development *Fund* (CP12/03080 and PI15/00071), by Chilean CONICYT/FONDECYT-POSTDOCTORADO N°3180486 (A.L.R.-C) and a fellowship from "La Caixa" Foundation (ID 100010434). The fellowship code is "LCF/BQ/DR19/11740016."

The content is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. National Institutes of Health or the Carlos III "Health Institute" (Spain).

References

Agarwal S, Zaman T, Murat Tuzcu E, Kapadia SR. 2011. Heavy metals and cardiovascular disease: results from the National Health and Nutrition Examination Survey (NHANES) 1999–2006. Angiology 62(5):422–429, PMID: 21421632, https://doi.org/10.1177/0003319710395562.

Akesson A, Bjellerup P, Lundh T, Lidfeldt J, Nerbrand C, Samsioe G, et al. 2006. Cadmium-induced effects on bone in a population-based study of women. Environ Health Perspect 114(6):830–834, PMID: 16759980, https://doi.org/10.1289/ehp.8763.

Andersson EM, Fagerberg B, Sallsten G, Borné Y, Hedblad B, Engström G, et al. 2018. Partial mediation by cadmium exposure of the association between tobacco smoking and atherosclerotic plaques in the carotid artery. Am J Epidemiol 187(4):806–816, PMID: 29020130, https://doi.org/10.1093/aje/kwx306.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. 2000. Gene ontology: tool for the unification of biology. Nat Genet 25(1):25–29, PMID: 10802651, https://doi.org/10.1038/75556.

ATSDR (Agency for Toxic Substances and Disease Registry). 2012. Toxicological profile: Cadmium. https://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=48&tid=15 [accessed 11 May 2020].

Benbrahim-Tallaa L, Waterland RA, Dill AL, Webber MM, Waalkes MP. 2007. Tumor suppressor gene inactivation during cadmium-induced malignant transformation of human prostate cells correlates with overexpression of de novo DNA methyltransferase. Environ Health Perspect 115(10):1454–1459, PMID: 17938735, https://doi.org/10.1289/ehp.10207.

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. 2010. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol 28(10):1045–1048, PMID: 20944595, https://doi.org/10.1038/nbt1010-1045.

Bojesen SE, Timpson N, Relton C, Davey Smith G, Nordestgaard BG. 2017. *AHRR* (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. Thorax 72(7):646–653, PMID: 28100713, https://doi.org/10.1136/thoraxjnl-2016-208789.

Bollepalli S, Korhonen T, Kaprio J, Anders S, Ollikainen M. 2019. EpiSmokEr: a robust classifier to determine smoking status from DNA methylation data. Epigenomics 11(13):1469–1486, PMID: 31466478, https://doi.org/10.2217/epi-2019-0206.

Breitling LP, Salzmann K, Rothenbacher D, Burwinkel B, Brenner H. 2012. Smoking, F2RL3 methylation, and prognosis in stable coronary heart disease. Eur Heart J 33(22):2841–2848, PMID: 22511653, https://doi.org/10.1093/eurheartj/ehs091.

Cassandri M, Smirnov A, Novelli F, Pitolli C, Agostini M, Malewicz M, et al. 2017. Zinc-finger proteins in health and disease. Cell Death Discov 3:17071, PMID: 29152378, https://doi.org/10.1038/cddiscovery.2017.71.

Chen K, Cheng Y, Berkout O, Lindhiem O. 2017. Analyzing proportion scores as outcomes for prevention trials: a statistical primer. Prev Sci 18(3):312–321, PMID: 26960687, https://doi.org/10.1007/s11121-016-0643-6.

Chen X, Zhu G, Jin T, Akesson A, Bergdahl IA, Lei L, et al. 2009. Changes in bone mineral density 10 years after marked reduction of cadmium exposure in a Chinese population. Environ Res 109(7):874–879, PMID: 19616207, https://doi.org/10.1016/j.envres.2009.06.003.

Chia N, Wang L, Lu X, Senut MC, Brenner C, Ruden DM. 2011. Hypothesis: environmental regulation of 5-hydroxymethylcytosine by oxidative stress. Epigenetics 6(7):853–856, PMID: 21617369, https://doi.org/10.4161/epi.6.7.16461.

Chiba M, Masironi R. 1992. Toxic and trace elements in tobacco and tobacco smoke. Bull World Health Organ 70(2):269–275, PMID: 1600587.

Chowdhury R, Ramond A, O'Keeffe LM, Shahzad S, Kunutsor SK, Muka T, et al. 2018. Environmental toxic metal contaminants and risk of cardiovascular disease: systematic review and meta-analysis. BMJ 362:k3310, PMID: 30158148, https://doi.org/10.1136/bmj.k3310.

de Smith AJ, Kaur M, Gonseth S, Endicott A, Selvin S, Zhang L, et al. 2017. Correlates of prenatal and early-life tobacco smoke exposure and frequency of common gene deletions in childhood acute lymphoblastic leukemia. Cancer Res 77(7):1674–1683, PMID: 28202519, https://doi.org/10.1158/0008-5472.CAN-16-2571.

Deen JF, Adams AK, Fretts A, Jolly S, Navas-Acien A, Devereux RB, et al. 2017. Cardiovascular disease in American Indian and Alaska native youth: unique risk factors and areas of scholarly need. J Am Heart Assoc 6: PMID: 29066451, https://doi.org/10.1161/JAHA.117.007576.

Delatte B, Jeschke J, Defrance M, Bachman M, Creppe C, Calonne E, et al. 2015. Genome-wide hydroxymethylcytosine pattern changes in response to oxidative stress. Sci Rep 5(1):12714, PMID: 26239807, https://doi.org/10.1038/srep12714.

Demanelis K, Virani S, Colacino JA, Basu N, Nishijo M, Ruangyuttikarn W, et al. 2017. Cadmium exposure and age-associated DNA methylation changes in non-smoking women from northern Thailand. Environ Epigenet 3:dvx006, PMID: 29492308, https://doi.org/10.1093/eep/dvx006.

Dempsie Y, MacRitchie NA, White K, Morecroft I, Wright AF, Nilsen M, et al. 2013. Dexfenfluramine and the oestrogen-metabolizing enzyme CYP1B1 in the development of pulmonary arterial hypertension. Cardiovasc Res 99(1):24–34, PMID: 23519266, https://doi.org/10.1093/cvr/cvt064.

Domingo-Relloso A, Grau-Perez M, Galan-Chilet I, Garrido-Martinez MJ, Tormos C, Navas-Acien A, et al. 2019. Urinary metals and metal mixtures and oxidative stress biomarkers in an adult population from Spain: the Hortega Study. Environ Int 123:171–180, PMID: 30529889, https://doi.org/10.1016/j.envint.2018.11.055.

Downward J. 2003. Targeting RAS signalling pathways in cancer therapy. Nat Rev Cancer 3(1):11–22, PMID: 12509763, https://doi.org/10.1038/nrc969.

Ernst J, Kellis M. 2017. Chromatin-state discovery and genome annotation with ChromHMM. Nat Protoc 12(12):2478–2492, PMID: 29120462, https://doi.org/10.1038/nprot.2017.124.

Everson TM, Punshon T, Jackson BP, Hao K, Lambertini L, Chen J, et al. 2018. Cadmium-associated differential methylation throughout the placental genome: epigenome-wide association study of two U.S. birth cohorts. Environ Health Perspect 126(1):017010, PMID: 29373860, https://doi.org/10.1289/EHP2192.

Fagerberg B, Borné Y, Barregard L, Sallsten G, Forsgard N, Hedblad B, et al. 2017. Cadmium exposure is associated with soluble urokinase plasminogen activator receptor, a circulating marker of inflammation and future cardiovascular disease. Environ Res 152:185–191, PMID: 27792942, https://doi.org/10.1016/j.envres.2016.10.019.

Fasanelli F, Baglietto L, Ponzi E, Guida F, Campanella G, Johansson M, et al. 2015. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nat Commun 6:10192, PMID: 26667048, https://doi.org/10.1038/ncomms10192.

Fortin JP, Triche TJ, Hansen KD, Hansen KD. 2017. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics 33(4):558–560, PMID: 28035024, https://doi.org/10.1093/bioinformatics/btw691.

Gao X, Jia M, Zhang Y, Breitling LP, Brenner H. 2015. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. Clin Epigenetics 7:113, PMID: 26478754, https://doi.org/10.1186/s13148-015-0148-3.

García-Esquinas E, Pollan M, Tellez-Plaza M, Francesconi KA, Goessler W, Guallar E, et al. 2014. Cadmium exposure and cancer mortality in a prospective cohort: the Strong Heart Study. Environ Health Perspect 122(4):363–370, PMID: 24531129, https://doi.org/10.1289/ehp.1306587.

Gonseth S, de Smith AJ, Roy R, Zhou M, Lee ST, Shao X, et al. 2016. Genetic contribution to variation in DNA methylation at maternal smoking-sensitive loci in exposed neonates. Epigenetics 11(9):664–673, PMID: 27403598, https://doi.org/10.1080/15592294.2016.1209614.

Grau-Perez M, Pichler G, Galan-Chilet I, Briongos-Figuero LS, Rentero-Garrido P, Lopez-Izquierdo R, et al. 2017. Urine cadmium levels and albuminuria in a general population from Spain: a gene-environment interaction analysis. Environ Int 106:27–36, PMID: 28558300, https://doi.org/10.1016/j.envint.2017.05.008.

Groth-Pedersen L, Aits S, Corcelle-Termeau E, Petersen NHT, Nylandsted J, Jäättelä M. 2012. Identification of cytoskeleton-associated proteins essential for lysosomal stability and survival of human cancer cells. PLoS One 7(10):e45381, PMID: 23071517, https://doi.org/10.1371/journal.pone.0045381.

Haarmann-Stemmann T, Bothe H, Kohli A, Sydlik U, Abel J, Fritsche E. 2007. Analysis of the transcriptional regulation and molecular function of the aryl hydrocarbon receptor repressor in human cell lines. Drug Metab Dispos 35(12):2262–2269, PMID: 17890447, https://doi.org/10.1124/dmd.107.016253.

Harlid S, Xu Z, Panduri V, Sandler DP, Taylor JA. 2014. CpG sites associated with cigarette smoking: analysis of epigenome-wide data from the Sister Study. Environ Health Perspect 122(7):673–678, PMID: 24704585, https://doi.org/10.1289/ehp.1307480.

Heuberger DM, Schuepbach RA. 2019. Protease-activated receptors (PARs): mechanisms of action and potential therapeutic modulators in PAR-driven inflammatory diseases. Thromb J 17:4, PMID: 30976204, https://doi.org/10.1186/s12959-019-0194-8.

Hossain MB, Vahter M, Concha G, Broberg K. 2012. Low-level environmental cadmium exposure is associated with DNA hypomethylation in Argentinean women.

Environ Health Perspect 120(6):879–884, PMID: 22382075, https://doi.org/10.1289/ehp.1104600.
Illumina Inc. 2019. Infinium MethylationEPIC Product Files. https://support.illumina.com/downloads/infinium-methylationepic-v1-0-product-files.html [accessed 9 May 2019].
Jiang G, Xu L, Song S, Zhu C, Wu Q, Zhang L, et al. 2008. Effects of long-term low-dose cadmium exposure on genomic DNA methylation in human embryo lung fibroblast cells. Toxicology 244(1):49–55, PMID: 18077075, https://doi.org/10.1016/j.tox.2007.10.028.
Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. 2016. Epigenetic signatures of cigarette smoking. Circ Cardiovasc Genet 9(5):436–447, PMID: 27651444, https://doi.org/10.1161/CIRCGENETICS.116.001506.
Joubert BR, Håberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, et al. 2012. 450K Epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environ Health Perspect 120(10):1425–1431, PMID: 22851337, https://doi.org/10.1289/ehp.1205412.
Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28(1):27–30, PMID: 10592173, https://doi.org/10.1093/nar/28.1.27.
Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. 2019. New approach for understanding genome variations in KEGG. Nucleic Acids Res 47(D1):D590–D595, PMID: 30321428, https://doi.org/10.1093/nar/gky962.
Lange T, Hansen JV. 2011. Direct and indirect effects in a survival context. Epidemiology 22(4):575–581, PMID: 21552129, https://doi.org/10.1097/EDE.0b013e31821c680c.
Larsson SC, Wolk A. 2016. Urinary cadmium and mortality from all causes, cancer and cardiovascular disease in the general population: systematic review and meta-analysis of cohort studies. Int J Epidemiol 45(3):782–791, PMID: 25997435, https://doi.org/10.1093/ije/dyv086.
Lee DH, Jacobs DR, Porta M. 2009. Hypothesis: a unifying mechanism for nutrition and chemicals as lifelong modulators of DNA hypomethylation. Environ Health Perspect 117(12):1799–1802, PMID: 20049195, https://doi.org/10.1289/ehp.0900741.
Lee ET, Welty TK, Fabsitz R, Cowan LD, Le NA, Oopik AJ, et al. 1990. The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. Am J Epidemiol 132(6):1141–1155, PMID: 2260546, https://doi.org/10.1093/oxfordjournals.aje.a115757.
Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28(6):882–883, PMID: 22257669, https://doi.org/10.1093/bioinformatics/bts034.
Leroy C, Landais E, Briault S, David A, Tassy O, Gruchy N, et al. 2013. The 2q37-deletion syndrome: an update of the clinical spectrum including overweight, brachydactyly and behavioural features in 14 new patients. Eur J Hum Genet 21(6):602–612, PMID: 23073310, https://doi.org/10.1038/ejhg.2012.230.
Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, et al. 2009. A new equation to estimate glomerular filtration rate. Ann Intern Med 150(9):604–612, PMID: 19414839, https://doi.org/10.7326/0003-4819-150-9-200905050-00006.
Li S, Wong EM, Bui M, Nguyen TL, Joo JE, Stone J, et al. 2018. Causal effect of smoking on DNA methylation in peripheral blood: a twin and family study. Clin Epigenetics 10:18, PMID: 29456763, https://doi.org/10.1186/s13148-018-0452-9.
McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. 2016. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. Genom Data 9:22–24, PMID: 27330998, https://doi.org/10.1016/j.gdata.2016.05.012.
McElroy JA, Kruse RL, Guthrie J, Gangnon RE, Robertson JD. 2017. Cadmium exposure and endometrial cancer risk: a large midwestern U.S. population-based case-control study. PLoS One 12(7):e0179360, PMID: 28742092, https://doi.org/10.1371/journal.pone.0179360.
Moitra S, Brashier BB, Sahu S. 2014. Occupational cadmium exposure-associated oxidative stress and erythrocyte fragility among jewelry workers in India. Am J Ind Med 57(9):1064–1072, PMID: 24812026, https://doi.org/10.1002/ajim.22336.
Morgan JC, Byron MJ, Baig SA, Stepanov I, Brewer NT. 2017. How people think about the chemicals in cigarette smoke: a systematic review. J Behav Med 40(4):553–564, PMID: 28224264, https://doi.org/10.1007/s10865-017-9823-5.
Möröy T, Khandanpour C. 2019. Role of GFI1 in epigenetic regulation of MDS and AML Pathogenesis: mechanisms and therapeutic implications. Front Oncol 9:824, PMID: 31508375, https://doi.org/10.3389/fonc.2019.00824.
Navas-Acien A, Selvin E, Sharrett AR, Calderon-Aranda E, Silbergeld E, Guallar E. 2004. Lead, cadmium, smoking, and increased risk of peripheral arterial disease. Circulation 109(25):3196–3201, PMID: 15184277, https://doi.org/10.1161/01.CIR.0000130848.18636.B2.
Navas-Acien A, Tellez-Plaza M, Guallar E, Muntner P, Silbergeld E, Jaar B, et al. 2009. Blood cadmium and lead and chronic kidney disease in US adults: a joint analysis. Am J Epidemiol 170(9):1156–1164, PMID: 19700501, https://doi.org/10.1093/aje/kwp248.
Nemmiche S. 2016. Oxidative signaling response to cadmium exposure. Toxicol Sci 156:4–10, PMID: 27803385, https://doi.org/10.1093/toxsci/kfw222.
Nica DV, Popescu C, Draghici GA, Andrica F-M, Privistirescu IA, Gergen II, et al. 2017. High-level dietary cadmium exposure is associated with global DNA hypermethylation in the gastropod hepatopancreas. PLoS One 12(9):e0184221, PMID: 28877233, https://doi.org/10.1371/journal.pone.0184221.
Niu L, Xu Z, Taylor JA. 2016. RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. Bioinformatics 32(17):2659–2663, PMID: 27153672, https://doi.org/10.1093/bioinformatics/btw285.
Nordberg GF, Nogawa K, Nordberg M. 2015. Chapter 32: Cadmium. In: *Handbook on the Toxicology of Metals*. Nordberg GF, Fowler BA, Nordberg M, eds. 4th ed. San Diego, CA: Academic Press, 667–716.
Olmedo P, Grau-Perez M, Fretts A, Tellez-Plaza M, Gil F, Yeh F, et al. 2017. Dietary determinants of cadmium exposure in the Strong Heart Family Study. Food Chem Toxicol 100:239–246, PMID: 28012896, https://doi.org/10.1016/j.fct.2016.12.015.
Orr SE, Bridges CC. 2017. Chronic kidney disease and exposure to nephrotoxic metals. Int J Mol Sci 18(5):1039, PMID: 28498320, https://doi.org/10.3390/ijms18051039.
Parmar P, Lowry E, Cugliari G, Suderman M, Wilson R, Karhunen V, et al. 2018. Association of maternal prenatal smoking GFI1-locus and cardio-metabolic phenotypes in 18,212 adults. EBioMedicine 38:206–216, PMID: 30442561, https://doi.org/10.1016/j.ebiom.2018.10.066.
Pearl J. 2012. The causal mediation formula–a guide to the assessment of pathways and mechanisms. Prev Sci 13(4):426–436, PMID: 22419385, https://doi.org/10.1007/s11121-011-0270-1.
Phipson B, Maksimovic J, Oshlack A. 2015. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics 32:btv560, PMID: 26424855, https://doi.org/10.1093/bioinformatics/btv560.
Poirier LA, Vlasova TI. 2002. The prospective role of abnormal methyl metabolism in cadmium toxicity. Environ Health Perspect 110(suppl 5):793–795, PMID: 12426133, https://doi.org/10.1289/ehp.02110s5793.
Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38(8):904–909, PMID: 16862161, https://doi.org/10.1038/ng1847.
Ravichandran M. 2017. Biochemical Characterisation of TET DNA Hydroxylases. https://elib.uni-stuttgart.de/bitstream/11682/9211/5/Biochemical%20Characterisation%20of%20TET%20DNA%20Hydroxylases-M.Ravichandran.pdf [accessed 14 March 2020].
Reynolds LM, Wan M, Ding J, Taylor JR, Lohman K, Su D, et al. 2015. DNA methylation of the aryl hydrocarbon receptor repressor associations with cigarette smoking and subclinical atherosclerosis. Circ Cardiovasc Genet 8(5):707–716, PMID: 26307030, https://doi.org/10.1161/CIRCGENETICS.115.001097.
Riancho J, Del Real A, Riancho JA. 2016. How to interpret epigenetic association studies: a guide for clinicians. Bonekey Rep 5:797, PMID: 27195108, https://doi.org/10.1038/bonekey.2016.24.
ROADMAP Epigenomics Project. Chromatine state learning. https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state [accessed 4 October 2019].
Ruiz-Hernandez A, Kuo C-C, Rentero-Garrido P, Tang W-Y, Redon J, Ordovas JM, et al. 2015. Environmental chemicals and DNA methylation in adults: a systematic review of the epidemiologic evidence. Clin Epigenetics 7:55, PMID: 25984247, https://doi.org/10.1186/s13148-015-0055-7.
Ruiz-Hernandez A, Navas-Acien A, Pastor-Barriuso R, Crainiceanu CM, Redon J, Guallar E, et al. 2017. Declining exposures to lead and cadmium contribute to explaining the reduction of cardiovascular mortality in the US population, 1988–2004. Int J Epidemiol 46(6):1903–1912, PMID: 29025072, https://doi.org/10.1093/ije/dyx176.
Sanders A, Smeester L, Rojas D, DeBussycher T, Wu M, Wright F, et al. 2014. Cadmium exposure and the epigenome: exposure-associated patterns of DNA methylation in leukocytes from mother-baby pairs. Epigenetics 9(2):212–221, PMID: 24169490, https://doi.org/10.4161/epi.26798.
Scheer J, Findenig S, Goessler W, Francesconi KA, Howard B, Umans JG, et al. 2012. Arsenic species and selected metals in human urine: validation of HPLC/ICPMS and ICPMS procedures for a long-term population-based epidemiological study. Anal Methods 4(2):406–413, PMID: 22685491, https://doi.org/10.1039/C2AY05638K.
Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–2504, PMID: 14597658, https://doi.org/10.1101/gr.1239303.
Shenker NS, Ueland PM, Polidoro S, van Veldhoven K, Ricceri F, Brown R, et al. 2013. DNA methylation as a long-term biomarker of exposure to tobacco smoke. Epidemiology 24(5):712–716, PMID: 23867811, https://doi.org/10.1097/EDE.0b013e31829d5cb3.
Sørensen KD, Abildgaard MO, Haldrup C, Ulhøi BP, Kristensen H, Strand S, et al. 2013. Prognostic significance of aberrantly silenced ANPEP expression in prostate cancer. Br J Cancer 108(2):420–428, PMID: 23322201, https://doi.org/10.1038/bjc.2012.549.
Srivastava RK, Pandey P, Rajpoot R, Rani A, Dubey RS. 2014. Cadmium and lead interactive effects on oxidative stress and antioxidative responses in rice

seedlings. Protoplasma 251(5):1047–1065, PMID: 24482190, https://doi.org/10.1007/s00709-014-0614-3.

Šrut M, Drechsel V, Höckner M. 2017. Low levels of Cd induce persisting epigenetic modifications and acclimation mechanisms in the earthworm Lumbricus terrestris. PLoS One 12(4):e0176047, PMID: 28426746, https://doi.org/10.1371/journal.pone.0176047.

Steenaard RV, Ligthart S, Stolk L, Peters MJ, van Meurs JB, Uitterlinden AG, et al. 2015. Tobacco smoking is associated with methylation of genes related to coronary artery disease. Clin Epigenet 7(1):54, PMID: 26015811, https://doi.org/10.1186/s13148-015-0088-y.

Stueve TR, Li W-Q, Shi J, Marconett CN, Zhang T, Yang C, et al. 2017. Epigenome-wide analysis of DNA methylation in lung tissue shows concordance with blood studies and identifies tobacco smoke-inducible enhancers. Hum Mol Genet 26(15):3014–3027, PMID: 28854564, https://doi.org/10.1093/hmg/ddx188.

Sun YV, Smith AK, Conneely KN, Chang Q, Li W, Lazarus A, et al. 2013. Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans. Hum Genet 132(9):1027–1037, PMID: 23657504, https://doi.org/10.1007/s00439-013-1311-6.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. 2019. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 47(D1):D607–D613, PMID: 30476243, https://doi.org/10.1093/nar/gky1131.

Takiguchi M, Achanzar WE, Qu W, Li G, Waalkes MP. 2003. Effects of cadmium on DNA-(Cytosine-5) methyltransferase activity and DNA methylation status during cadmium-induced cellular transformation. Exp Cell Res 286(2):355–365, PMID: 12749863, https://doi.org/10.1016/s0014-4827(03)00062-4.

Tellez-Plaza M, Guallar E, Howard BV, Umans JG, Francesconi KA, Goessler W, et al. 2013a. Cadmium exposure and incident cardiovascular disease. Epidemiology 24(3):421–429, PMID: 23514838, https://doi.org/10.1097/EDE.0b013e31828b0631.

Tellez-Plaza M, Jones MR, Dominguez-Lucas A, Guallar E, Navas-Acien A. 2013b. Cadmium exposure and clinical cardiovascular disease: a systematic review. Curr Atheroscler Rep 15(10):356, PMID: 23955722, https://doi.org/10.1007/s11883-013-0356-2.

Tellez-Plaza M, Navas-Acien A, Caldwell KL, Menke A, Muntner P, Guallar E. 2012. Reduction in cadmium exposure in the United States population, 1988–2008: the contribution of declining smoking rates. Environ Health Perspect 120(2):204–209, PMID: 22062584, https://doi.org/10.1289/ehp.1104020.

Teschendorff AE, Relton CL. 2018. Statistical and integrative system-level analysis of DNA methylation data. Nat Rev Genet 19(3):129–147, PMID: 29129922, https://doi.org/10.1038/nrg.2017.86.

The Gene Ontology Consortium. 2018. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res 47:D330–D338, PMID: 30395331, https://doi.org/10.1093/nar/gky1055.

Triche TJ, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. 2013. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res 41(7):e90–e90, PMID: 23476028, https://doi.org/10.1093/nar/gkt090.

Tsaprouni LG, Yang T-P, Bell J, Dick KJ, Kanoni S, Nisbet J, et al. 2014. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. Epigenetics 9(10):1382–1396, PMID: 25424692, https://doi.org/10.4161/15592294.2014.969637.

U.S. Department of Health and Human Services. 2010. How Tobacco Smoke Causes Disease: the Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Valeri L, Vanderweele TJ. 2013. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods 18(2):137–150, PMID: 23379553, https://doi.org/10.1037/a0031034.

van Iterson M, van Zwet EW, Bios Consortium, Heijmans BT. 2017. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. Genome Biol 18(1):19, PMID: 28129774, https://doi.org/10.1186/s13059-016-1131-9.

Vidal AC, Semenova V, Darrah T, Vengosh A, Huang Z, King K, et al. 2015. Maternal cadmium, iron and zinc levels, DNA methylation and birth weight. BMC Pharmacol Toxicol 16(1):20, https://doi.org/10.1186/s40360-015-0020-2.

Virani S, Rentschler KM, Nishijo M, Ruangyuttikarn W, Swaddiwudhipong W, Basu N, et al. 2016. DNA methylation is differentially associated with environmental cadmium exposure based on sex and smoking status. Chemosphere 145:284–290, PMID: 26688266, https://doi.org/10.1016/j.chemosphere.2015.10.123.

Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, et al. 2012. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. Hum Mol Genet 21(13):3073–3082, PMID: 22492999, https://doi.org/10.1093/hmg/dds135.

Wang X, Dasari S, Nowakowski GS, Lazaridis KN, Wieben ED, Kadin ME, et al. 2017. Retinoic acid receptor alpha drives cell cycle progression and is associated with increased sensitivity to retinoids in T-cell lymphoma. Oncotarget 8(16):26245–26255, PMID: 28412739, https://doi.org/10.18632/oncotarget.15441.

Warton DI, Hui F. 2011. The arcsine is asinine: the analysis of proportions in ecology. Ecology 92(1):3–10, PMID: 21560670, https://doi.org/10.1890/10-0340.1.

Wiklund P, Karhunen V, Richmond RC, Parmar P, Rodriguez A, De Silva M, et al. 2019. DNA methylation links prenatal smoking exposure to later life health outcomes in offspring. Clin Epigenet 11(1):97, PMID: 31262328, https://doi.org/10.1186/s13148-019-0683-4.

Xiong J, Liu X, Cheng QY, Xiao S, Xia LX, Yuan BF, et al. 2017. Heavy metals induce decline of derivatives of 5-methycytosine in both DNA and RNA of stem cells. ACS Chem Biol 12(6):1636–1643, PMID: 28448110, https://doi.org/10.1021/acschembio.7b00170.

Yao R, Han D, Sun X, Xie Y, Wu Q, Fu C, et al. 2018. Scriptaid inhibits cell survival, cell cycle, and promotes apoptosis in multiple myeloma via epigenetic regulation of p21. Exp Hematol 60:63–72, PMID: 29305109, https://doi.org/10.1016/j.exphem.2017.12.012.

Yuan D, Ye S, Pan Y, Bao Y, Chen H, Shao C. 2013. Long-term cadmium exposure leads to the enhancement of lymphocyte proliferation via down-regulating p16 by DNA hypermethylation. Mutat Res 757(2):125–131, PMID: 23948183, https://doi.org/10.1016/j.mrgentox.2013.07.007.

Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. 2013. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One 8(5):e63812, PMID: 23691101, https://doi.org/10.1371/journal.pone.0063812.

Zhang C, Liang Y, Lei L, Zhu G, Chen X, Jin T, et al. 2013a. Hypermethylations of RASAL1 and KLOTHO is associated with renal dysfunction in a Chinese population environmentally exposed to cadmium. Toxicol Appl Pharmacol 271(1):78–85, PMID: 23665422, https://doi.org/10.1016/j.taap.2013.04.025.

Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H. 2013b. F2RL3 methylation as a biomarker of current and lifetime smoking exposures. Environ Health Perspect, PMID: 24273234, https://doi.org/10.1289/ehp.1306937.

Zhang Y, Yang R, Burwinkel B, Breitling LP, Holleczek B, Schöttker B, et al. 2014. F2RL3 methylation in blood DNA is a strong predictor of mortality. Int J Epidemiol 43(4):1215–1225, PMID: 24510982, https://doi.org/10.1093/ije/dyu006.

Zheng LY, Umans JG, Yeh F, Francesconi KA, Goessler W, Silbergeld EK, et al. 2015. The association of urine arsenic with prevalent and incident chronic kidney disease: evidence from the Strong Heart Study. Epidemiology 26(4):601–612, PMID: 25929811, https://doi.org/10.1097/EDE.0000000000000313.

Zhu S, Bjorge JD, Fujita DJ. 2007. PTP1B Contributes to the oncogenic properties of colon cancer cells through Src activation. Cancer Res 67(21):10129–10137, PMID: 17974954, https://doi.org/10.1158/0008-5472.CAN-06-4338.